ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 979-8-3503-6874-1/25/$31.00 ©2025 IEEE | DOI: 10.1109/ICASSP49660.2025.10888974

Spectral Low-Rank Attention with Flow-Based Refinement for Spectral Reconstruction

Yiwen Wang
School of Computing and Artificial Intelligence
Southwestern University of Finance and Economics
Chengdu, China
15530848979@163.com

Zixin Tang
School of Computing and Artificial Intelligence
Southwestern University of Finance and Economics
Chengdu, China
sctangzixing123@163.com

Yexun Hu*
School of Computing and Artificial Intelligence
Southwestern University of Finance and Economics
Chengdu, China
huyexun2022@163.com

Guisong Liu
School of Computing and Artificial Intelligence
Southwestern University of Finance and Economics
Chengdu, China
gliu@swufe.edu.cn

Tai-Xiang Jiang
School of Computing and Artificial Intelligence
Southwestern University of Finance and Economics
Chengdu, China
taixiangjiang@gmail.com

***Abstract*—Spectral super-resolution (SSR) from RGB images, which involves reconstructing hyperspectral images (HSIs) from color images, has recently received great attention. While convolutional neural network (CNN)-based methods have demonstrated strong performance, they often overlook the self-similarity across the spectral dimension of HSIs. Transformer-based approaches have addressed this limitation by leveraging self-attention mechanisms to capture spectral correlations. However, these methods encounter computational and memory overheads that scale quadratically with the size of the HSIs. To overcome these challenges, we introduce a novel Spectral-wise LOw-Rank Attention (SLORA) mechanism that captures inter-spectral consistency in a low-dimensional space, thereby reducing both computational costs and model complexity. Additionally, we propose a flow-based refinement module to enhance generalization and performance on unseen HSIs. Experimental results from the NTIRE 2022 spectral reconstruction challenge and the spectral snapshot compression imaging task datasets validate the superiority of our method over state-of-the-art approaches.**

***Index Terms*—Attention, transformer, spectral super-resolution, spectral snapshot compression imaging.**

I. Introduction

Hyperspectral images (HSIs) capture extensive spectral information from real-world scenes, valuable for applications like medical imaging, remote sensing, and object tracking [1]–[3]. However, acquiring HSIs is time-consuming, particularly in dynamic scenarios where balancing spatial and spectral resolutions is challenging. Spectral snapshot compression imaging (SSCI) techniques like CASSI [4] compress spectral data into a coded matrix but are costly and prone to information loss [5]–[7]. Spectral super-resolution (SSR) methods aim to reconstruct HSIs from RGB images without hardware constraints but are limited by the ill-posed nature of the problem, relying on hand-designed priors that often fall short in real-world scenarios.

With the development of deep learning, many CNN-based methods [8], [9] have shown promise in HSI reconstruction. However, because the same substances have identical spectral response characteristics, also reflected in the unity of spectral vectors and the low-dimensional spectral structure of HSIs, which had been characterized by utilizing low-rank prior in many HSI-relevant methods [10], [11], these CNN-based methods are insufficient to portray the inter-spectral self-similarity and have limited ability to capture the long-range dependence of HSI [12]. Recently, Transformer [13] has been widely used in many fields due to the excellent ability of the self-attention mechanism to characterize long-range dependencies and complex features of data. Some SSR methods [12], [14] have gained even better performances by introducing the transformer structure. The core idea of self-attention is calculating the affinity between features to capture long-range dependencies, but as the size of the feature map increases, the computing and memory overheads increase quadratically. Additionally, self-attention only considers the

This work is supported in part by the Sichuan Science and Technology Program under Grant 2024ZYD0147, in part by the Natural Science Foundation of Xinjiang Uygur Autonomous Region under Grant 2024D01A18, in part by the National Natural Science Foundation of China (NSFC) under Grant 62376228, in part by Chengdu Science and Technology program under Grant 2023 JB00-00016-GX, in part by the Guanghua Talent Project., and in part by the Graduate Representative Achievement Cultivation Project of Southwest University of Finance and Economics under Grant JGS2024069.

* Corresponding Author

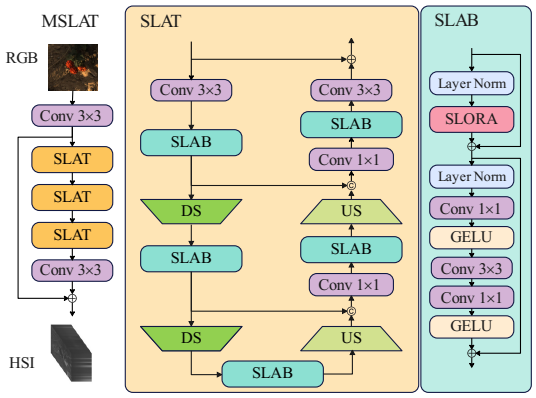
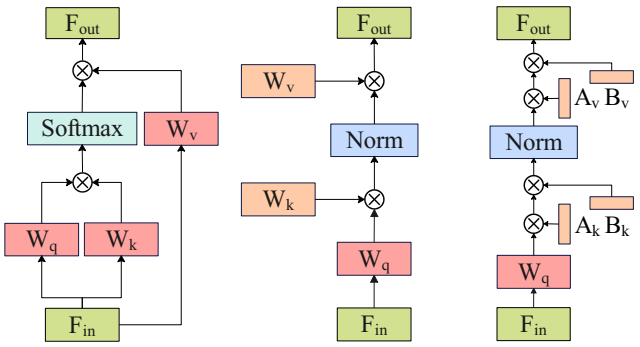


Fig. 1. Architecture of our MSLAT, which follows the architecture of MST++ [12].



(a) Self-Attention [13] (b) External-Attention [15] (c) SLORA

Fig. 2. Different attention mechanisms

relation between elements within a data sample and ignores potential relationships between elements in different samples, potentially limiting the ability and flexibility of self-attention [15], which implies Transformer-based SSR methods overlook the association among spectral vectors in different HSIs.

To address these challenges, we build on the MST++ framework [12] and introduce Spectral-wise LOw-Rank Attention (SLORA), which focuses on the low-dimensional structure of inter-spectral consistency in attention keys and values, rather than directly modeling inter-spectral relations. As shown in Fig. 1, our Multi-stage Spectral Low-rAnk Transformer (MSLAT) follows the MST++ architecture, replacing the key module with the proposed SLAT block that incorporates SLORA. SLORA extends extension-attention [15] by using global memory units ($\mathbf{W}_k$, $\mathbf{W}_v$, Fig. 2b) shared across samples, with low-rank factor components added for efficiency (Fig. 2c), similar to LoRA [16]. This significantly reduces storage and computational costs while capturing inter-spectral relationships. To address the potential loss of high-frequency spatial details, we introduce a flow-based refinement module (FRM) within a flow-based learning framework [17]–[20]. During refinement, the SLAT-based encoder parameters are frozen, using its output as pseudo-labels, while the refinement module, trained with a Gaussian distribution, maps input HSI images into latent space to learn the conditional distribution of clean HSIs.

Our contributions are as follows:

- We propose SLORA, a spectral-wise low-rank attention mechanism that efficiently captures inter-spectral relationships in HSIs while reducing computational costs and parameters. Additionally, we introduce a flow-based refinement module (FRM) that enhances network generalization and performance on unseen HSIs, acting as a plug-and-play tool during test-time training.
- Our method outperforms state-of-the-art approaches in the NTIRE 2022 Spectral Reconstruction Challenge [21] with fewer parameters and FLOPS in the SSCI task.

II. Method

First, as illustrated in Fig. 1, our Multi-stage Spectral Low-rAnk Transformer (MSLAT) takes RGB images as input and outputs the corresponding HSIs. The architecture comprises three cascaded Spectral Low-rAnk Transformer (SLAT) blocks. Initially, a 3-channel RGB image passes through a 3×3 convolutional layer to expand the output feature map to 31 channels, which is then fed into the SLAT blocks.

Each SLAT block features a U-shaped structure [22] that captures multi-scale spatial features while reducing computational costs through upsampling and downsampling. It consists of two 3×3 convolutional layers at the beginning and end, two 4×4 convolutional layers (stride = 2) for downsampling (DS), two 2×2 transposed convolutional layers (stride = 2) for upsampling (US), two 1×1 convolutional layers, and several Spectral Low-rAnk Blocks (SLAB) to model the low-dimensional structure of inter-spectral consistency. Additionally, three residual connections are incorporated to address spatial information loss during downsampling and enhance stability.

To efficiently capture inter-spectral relationships in hyperspectral images (HSIs) while reducing computational and memory costs, we propose a novel attention mechanism called Spectral-wise LOw-Rank Attention (SLORA), as shown in Fig. 2. In SLORA, the key and value memory units ($\mathbf{W}_k$, $\mathbf{W}_v$) from the entire training dataset, as used in external-attention [15], are factorized into low-rank components: $\mathbf{A}_k$, $\mathbf{B}_k$ and $\mathbf{A}_v$, $\mathbf{B}_v$. Given an input feature map $\mathbf{F}_{\text{in}} \in \mathbb{R}^{n \times d}$, where n is the sequence length and d is the representation dimension, SLORA can be expressed as:

$$\mathbf{F} = \mathbf{F}_{\text{in}}\mathbf{W}_q, \ \mathbf{M} = \text{Norm}(\mathbf{F}\mathbf{A}_k\mathbf{B}_k), \ \mathbf{F}_{\text{out}} = \mathbf{M}\mathbf{A}_v\mathbf{B}_v,$$

where $\mathbf{W}_q \in \mathbb{R}^{d \times d}$ is the query parameter matrix, $\mathbf{A}_k \in \mathbb{R}^{d \times r}$ and $\mathbf{B}_k \in \mathbb{R}^{r \times s}$ are the low-rank components for the key, and $\mathbf{A}_v \in \mathbb{R}^{s \times r}$ and $\mathbf{B}_v \in \mathbb{R}^{r \times d}$ are for the value, with $r < \min(s, d)$. The attention map is calculated by the affinities between query vectors and the low-rank key memory ($\mathbf{A}_k\mathbf{B}_k$), followed by generating a refined feature map using the low-rank value memory ($\mathbf{A}_v\mathbf{B}_v$). This approach effectively captures the low-dimensional structure of inter-spectral relationships among various spectral vectors in HSIs

TABLE I

QUANTITATIVE RESULTS BY DIFFERENT METHODS ON NTIRE 2022 SPECTRAL CHALLENGE (VALID). THE BEST AND THE SECOND BEST VALUES ARE RESPECTIVELY HIGHLIGHTED BY RED AND BLUE COLORS.

Method	MRAE↓	RMSE↓	PSNR↑	Params(M)↓	FLOPs(G)↓
AWAN [23]	0.2500	0.0367	31.22	4.04	270.61
HDNet [24]	0.2048	0.0317	32.13	2.66	173.81
HINet [25]	0.2032	0.0303	32.51	5.21	31.04
MIRNET [26]	0.1890	0.0274	33.29	3.75	42.95
Restormer [14]	0.1833	0.0274	33.40	15.11	93.77
MPRNET [27]	0.1817	0.0270	33.50	3.62	101.59
MST++ [12]	0.1645	0.0248	34.32	1.62	23.05
MSLAT	0.1620	0.0247	34.11	1.54	19.89
RMSLAT	0.1567	0.0245	34.68	1.85	40.21

TABLE II

QUANTITATIVE RESULTS BY DIFFERENT METHODS ON 10 SCENES OF CAVE SIMULATION DATASET. PSNR AND SSIM, PARAMS, FLOPS ARE REPORTED. THE BEST AND THE SECOND BEST VALUES ARE RESPECTIVELY HIGHLIGHTED BY RED AND BLUE COLORS.

Method	PSNR↑	SSIM ↓	Params(M)↓	FLOPs(G)↓
λ-Net [28]	28.53	0.841	62.64	117.98
DGSMP [29]	32.63	0.917	3.76	646.65
HDNet [24]	34.97	0.943	2.37	154.76
CST-L [30]	36.12	0.957	3.00	40.11
BIRNAT [31]	37.58	0.960	4.40	2122.66
DAUHST-9stg [32]	38.36	0.967	6.15	79.50
MSLAT	36.44	0.963	2.03	26.81
RMSLAT	37.12	0.965	2.34	46.76

while reducing memory usage and computational complexity. The complexity of SLORA ($O(nr(d+s))$) is lower than that of self-attention ($O(n^2d)$). The multi-head mechanism is omitted here for simplicity.

After training MSLAT, a flow-based refinement module (FRM) is introduced to combine with MSLAT, constituting a Refined Multi-stage Spectral Low-rAnk Transformer (RMSLAT), to enhance generalization and accuracy. This module, FRM, inspired by the HIDFlowNet decoder, learns the conditional distribution of hyperspectral images rather than a deterministic mapping, its invertible design allows it to map inputs into a latent space following a simple distribution (like a Gaussian distribution) and recover them accurately. With the employment of FRM, the refined output is ensured better alignment with the expected distribution. For further details on the architecture, please refer to HIDFlowNet [20].

III. EXPERIMENTS

a) Datasets: Two dominant hyperspectral image generation tasks: spectral super-resolution from RGB (RGB2HSI) and spectral snapshot compression imaging (SSCI), are considered to validate the performance of the proposed method. For RGB2HSI, we adopt NTIRE 2022 Spectral Reconstruction Challenge dataset [21], which is divided into 900 training RGB-HSI pairs, 50 validation RGB-HSI pairs, and 50 RGB-HSI pairs for test (unavailable). Each HSI has a spatial resolution of 482×512 and 31 spectral bands from 400nm to 700nm. Since HSIs of the test set are not available, we employ the validation set for evaluation.

For SSCI, we performed simulations using two publicly available spectral datasets: CAVE [33] and KAIST [34]. The CAVE dataset consists of 32 HSIs with a spatial size of 512×512 and 31 spectral bands. The KAIST dataset includes 30 HSIs with a spatial size of 2704×3376 with 31 spectral bands. Following the experimental settings in previous works [35]–[37], We used the CAVE dataset as a training set and then extracted 10 scenes with a spatial size of 256×256 from KAIST as a test set for testing. To match the wavelength range of the real system, we selected 28 bands from 450 nm to 650 nm for the training and test data by spectral interpolation.

b) Implementation Details: We implemented the proposed method MSLAT and RMSLAT in Pytorch and trained the model with the Adam optimizer. The model was trained 400 times on a single RTX 3090 GPU using the Cosine Annealing scheme. The initial learning rate was set to 4×10^{-4}. The batch size was set to 20. It is worth noting that the rank r in SLORA is set to $r = \lfloor c * 0.8 \rfloor$, where c is the number of channels of the input feature. During the training of MSLAT, MRAE loss function L_1 between the Groundtruth (GT) and the output of MSLAT is employed, while during the refinement, we freeze the parameter of MSLAT and plug FRM onto it, then train FRM, denoted as f_θ, by minimizing the loss L_{re} defined as

$$L_{re}(\mathcal{X}, \mathcal{Y}, \mathcal{Z}) = \lambda_1 L_1(\mathcal{X}, f_\theta^{-1}(\mathcal{Z}; \mathcal{Y})) + \lambda_2 L_{nll}(\mathcal{X}, \mathcal{Y}), \quad (1)$$

where λ_1, λ_2 are nonnegative parameters, f_θ^{-1} is the inverse transforms of f_θ, and L_{nll} is the negative log-likelihood (NLL) loss in [19]. When the test-time training is complete, the final output is a weighted sum of the output from MSLAT and FRM, with weighting factors of 0.8 and 0.2, respectively.

A. Comparisons with State-of-the-Art Methods

For RGB2HSI, we compared the proposed method with various methods including [12], [14], [24]–[27], and the results are shown in Table 1. Among them, AWAN is the winner of the clean track of the NTIRE 2020 Spectral super-resolution Challenge.HR Net is the winner of the real-world track of the NTIRE 2020 Spectral super-resolution Challenge.MST++ is the first-place winner of the NTIRE 2022 Spectral super-resolution Challenge. Restormer and MPRNet are advanced network architectures in the field of image restoration. Following the NTIRE 2022 Spectral Reconstruction Challenge setup, the metrics are calculated only within the central 226 × 256 region of each image. It is clear that the proposed method achieves state-of-the-art performance with the lowest error and computational cost.

For SSCI task, different methods including [24], [28]–[32] are selected for comparison. Although RMSLAT does not outperform DAUHST-9stg [32] and BIRNAT [31], obviously the proposed method has far fewer parameters and computational complexity, which demonstrates that the proposed method better balances performance and computational effort.

B. Ablation Study

To verify the effectiveness and impact of two components, SLORA and FRM, adopted in this work, we first compare MSLAT (integrating SLORA) with MST++ [12] (utilizing

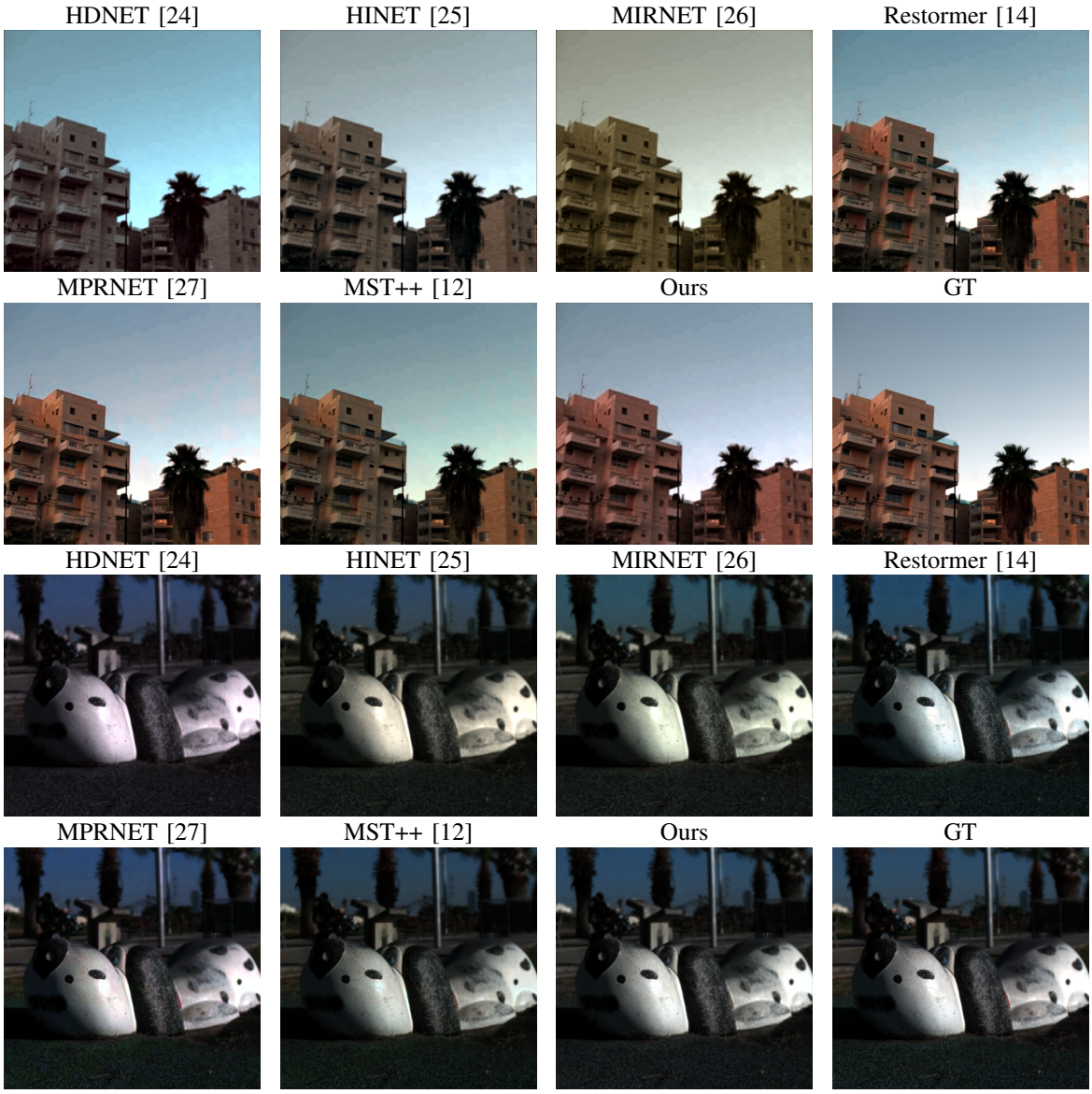


Fig. 3. Pseudo color images (constructed by three bands: R-25th G-15th B-10th) of results by different methods on HSIs (#901 and #907 in the NTIRE 2022 valid dataset).

TABLE III
ABLATION STUDY OF SLORA AND FRM. THE BEST VALUES ARE RESPECTIVELY HIGHLIGHTED BY RED COLOR.

Attention	MST++	MSLAT	RMST++	RMSLAT
MRAE↑	0.1645	0.1620	0.1614	0.1567
PSNR↑	34.32	34.54	34.44	34.68
Params↓	1.62	1.54	1.93	1.85
FLOPs↓	22.29	19.89	42.97	39.84

self-attention [13]), with both models trained on the RGB2HSI task with the same scheme. We then examine the effect of combining these two models with FRM. As shown in Table III, it is obvious that the SLORA-based model (MSLAT) outperforms the SA-based model (MST++) with lower memory requirements and computational complexity, demonstrating the feasibility of SLORA. Moreover, introducing the FRM into these models clearly enhances their performance, and the proposed method (RMSLAT) surpasses the MST++ combined with the FRM (RMST++), thus proving the flexibility and efficacy of FRM.

CONCLUSION

In this work, we focus on spectral reconstruction tasks and propose a novel spectral-wise low-rank attention (SLORA) mechanism to depict inter-spectral consistency among various HSIs with lower computational complexity and memory overheads. In addition, a flow-based refinement module (FRM) is employed to further boost the generalization and performance. Experiments compared with SOTA methods on the dataset of NTIRE 2022 Spectral Reconstruction Challenge and SSCI demonstrate the superiority of the proposed method.

REFERENCES

[1] V. Backman, M. B. Wallace, L. Perelman, J. Arendt, R. Gurjar, M. Müller, Q. Zhang, G. Zonios, E. Kline, T. McGillican, *et al.*, "Detection of preinvasive cancer cells," *Nature*, vol. 406, no. 6791, pp. 35–36, 2000.

[2] M. Borengasser, W. S. Hungate, and R. Watkins, *Hyperspectral remote sensing: principles and applications*. CRC press, 2007.

[3] J. Li, C. Wu, R. Song, Y. Li, and F. Liu, "Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from rgb images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 462–463, 2020.

[4] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Applied Optics*, p. B44, Apr 2008.

[5] G. R. Arce, D. J. Brady, L. Carin, H. Arguello, and D. S. Kittle, "Compressive coded aperture spectral imaging: An introduction," *IEEE Signal Processing Magazine*, vol. 31, no. 1, pp. 105–115, 2013.

[6] L. Wang, Z. Xiong, D. Gao, G. Shi, and F. Wu, "Dual-camera design for coded aperture snapshot spectral imaging," *Applied optics*, vol. 54, no. 4, pp. 848–858, 2015.

[7] H. Arguello and G. R. Arce, "Rank minimization code aperture design for spectrally selective compressive imaging," *IEEE transactions on image processing*, vol. 22, no. 3, pp. 941–954, 2012.

[8] Y. Zhao, L.-M. Po, Q. Yan, W. Liu, and T. Lin, "Hierarchical regression network for spectral reconstruction from rgb images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 422–423, 2020.

[9] R. Delgado-Escano, F. M. Castro, J. R. Cozar, M. J. Marin-Jimenez, and N. Guil, "An end-to-end multi-task and fusion cnn for inertial-based gait recognition," *IEEE Access*, vol. 7, pp. 1897–1908, 2018.

[10] L. Zhuang, L. Gao, B. Zhang, X. Fu, and J. M. Bioucas-Dias, "Hyperspectral image denoising and anomaly detection based on low-rank and sparse representations," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–17, 2020.

[11] T.-X. Jiang, X.-L. Zhao, H. Zhang, and M. K. Ng, "Dictionary learning with low-rank coding coefficients for tensor completion," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 2, pp. 932–946, 2023.

[12] Y. Cai, J. Lin, Z. Lin, H. Wang, Y. Zhang, H. Pfister, R. Timofte, and L. Van Gool, "Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 745–755, 2022.

[13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," *Neural Information Processing Systems,Neural Information Processing Systems*, Jun 2017.

[14] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M.-H. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5728–5739, 2022.

[15] M.-H. Guo, Z.-N. Liu, T.-J. Mu, and S.-M. Hu, "Beyond self-attention: External attention using two linear layers for visual tasks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 5436–5447, 2022.

[16] E. J. Hu, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen, *et al.*, "Lora: Low-rank adaptation of large language models," in *International Conference on Learning Representations*.

[17] X. Han, X. Hu, W. Huang, and M. R. Scott, "Clothflow: A flow-based model for clothed person generation," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10471–10480, 2019.

[18] H. Sun, R. Mehta, H. H. Zhou, Z. Huang, S. C. Johnson, V. Prabhakaran, and V. Singh, "Dual-glow: Conditional flow-based generative model for modality transfer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10611–10620, 2019.

[19] A. Lugmayr, M. Danelljan, L. Van Gool, and R. Timofte, "Srflow: Learning the super-resolution space with normalizing flow," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V 16*, pp. 715–732, Springer, 2020.

[20] L. Pang, W. Gu, X. Cao, X. Rui, J. Peng, S. Xu, G. Yang, and D. Meng, "Hidflownet: A flow-based deep network for hyperspectral image denoising," *arXiv preprint arXiv:2306.17797*, 2023.

[21] B. Arad, R. Timofte, R. Yahel, N. Morag, A. Bernat, Y. Cai, J. Lin, Z. Lin, H. Wang, Y. Zhang, *et al.*, "Ntire 2022 spectral recovery challenge and data set," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 863–881, 2022.

[22] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pp. 234–241, Springer, 2015.

[23] J. Li, C. Wu, R. Song, Y. Li, and F. Liu, "Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from rgb images," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pp. 462–463, 2020.

[24] X. Hu, Y. Cai, J. Lin, H. Wang, X. Yuan, Y. Zhang, R. Timofte, and L. Van Gool, "Hdnet: High-resolution dual-domain learning for spectral compressive imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 17542–17551, 2022.

[25] L. Chen, X. Lu, J. Zhang, X. Chu, and C. Chen, "Hinet: Half instance normalization network for image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 182–192, 2021.

[26] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Learning enriched features for real image restoration and enhancement," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pp. 492–511, Springer, 2020.

[27] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao, "Multi-stage progressive image restoration," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14821–14831, 2021.

[28] X. Miao, X. Yuan, Y. Pu, and V. Athitsos, "l-net: Reconstruct hyperspectral images from a snapshot measurement," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2019.

[29] T. Huang, W. Dong, X. Yuan, J. Wu, and G. Shi, "Deep gaussian scale mixture prior for spectral compressive imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16216–16225, June 2021.

[30] Y. Cai, J. Lin, X. Hu, H. Wang, X. Yuan, Y. Zhang, R. Timofte, and L. Van Gool, "Coarse-to-fine sparse transformer for hyperspectral image reconstruction," in *Computer Vision – ECCV 2022* (S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, eds.), (Cham), pp. 686–704, Springer Nature Switzerland, 2022.

[31] Z. Cheng, B. Chen, R. Lu, Z. Wang, H. Zhang, Z. Meng, and X. Yuan, "Recurrent neural networks for snapshot compressive imaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 2264–2281, 2023.

[32] Y. Cai, J. Lin, H. Wang, X. Yuan, H. Ding, Y. Zhang, R. Timofte, and L. V. Gool, "Degradation-aware unfolding half-shuffle transformer for spectral compressive imaging," in *Advances in Neural Information Processing Systems* (S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, eds.), vol. 35, pp. 37749–37761, Curran Associates, Inc., 2022.

[33] J.-I. Park, M.-H. Lee, M. D. Grossberg, and S. K. Nayar, "Multispectral imaging using multiplexed illumination," in *2007 IEEE 11th International Conference on Computer Vision*, pp. 1–8, IEEE, 2007.

[34] I. Choi, M. Kim, D. Gutierrez, D. Jeon, and G. Nam, "High-quality hyperspectral reconstruction using a spectral prior," tech. rep., 2017.

[35] Z. Meng, J. Ma, and X. Yuan, "End-to-end low cost compressive spectral imaging with spatial-spectral self-attention," in *European conference on computer vision*, pp. 187–204, Springer, 2020.

[36] T. Huang, W. Dong, X. Yuan, J. Wu, and G. Shi, "Deep gaussian scale mixture prior for spectral compressive imaging," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16216–16225, 2021.

[37] Y. Cai, J. Lin, X. Hu, H. Wang, X. Yuan, Y. Zhang, R. Timofte, and L. Van Gool, "Coarse-to-fine sparse transformer for hyperspectral image reconstruction," in *European conference on computer vision*, pp. 686–704, Springer, 2022.